

The assembled EST sequence of the novel protein is confirmed by a peptide mass fingerprinting and RACE. The coding region of hcc-1 cDNA has 630 bases, which code for the 210 amino acids of the full-length protein. The unique DNA sequence at the 3' untranslated region (218 bp) has been used to localize the gene to chromosome 7q22.1. A total of 690 bp at the 5' untranslated region of hcc-1 has been identified and promoter activity has been demonstrated at this region. A number of uORFs, which is a common feature in proto-oncogenes and growth factors, are noted at the 5' untranslated region.--

D<sub>1</sub> --[0009] The protein HCC-1 is localized to the nucleus region of two liver cell lines by immunofluorescence staining. Bioinformatics predictions show that the first 42 amino acids of the protein have identity matches to heterogenous nuclear ribonucleoproteins from various vertebrate species including human. The domain is also a putative bi-helical DNA-binding motif. The rest of the HCC-1 amino acid sequence has no known homology in vertebrates.--

D<sub>2</sub> --[0027] The term "differential" or a related term such as "differentially" in relation to gene expression means that a gene sequence is expressed in one type of cell or tissue (e.g. cancerous cell or tissue) but is substantially not expressed in another cell or tissue. The term "preferential" or a related term such as "preferentially" in relation to gene expression means that a gene sequence is expressed at a higher level in one type of cell or tissue (e.g. cancerous cell or tissue) relative to another type of cell or tissue. The difference in expression levels may, for example, be from two-fold to 100-fold or from three-fold to 50-fold. In one embodiment, the gene is liver tissue of patients within hepatocellular carcinoma and is substantially not expressed in the normal liver.--

D<sub>3</sub> --[0064] The binding processes are well-known in the art and generally consist of cross-linking, covalently binding, or physically adsorbing. The polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from room temperature to 25°C or above) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.--

D<sup>3</sup>  
1

--[0102] The first dimensional IEF was performed on precast 18 cm IPG strips (Amersham Pharmacia Biotech) at 20°C with a maximum current setting of 50  $\mu$ A/strip using an Amersham Pharmacia IPGphor IEF unit. The strips were rehydrated for a minimum of 10 hrs in ceramic strip holders in 350 FL of sample containing 7 M urea, 2 M thiourea, 4% v/v CHAPS, 1 mM PMSF, 20 mM dithiothreitol (DTT) (Bio-Rad) and 0.5% v/v IPG buffer (Amersham Pharmacia Biotech). The amount of protein loaded was ~150  $\mu$ g for analytical gels and ~400  $\mu$ g protein for preparative gels. A low voltage of 30 V was applied during rehydration. After rehydration, IEF run was carried out using the following conditions: (i) 500 V, 500 Vhr; (ii) 1,000 V, 1000 Vhr; and (iii) 8000 V, 32000 Vhr. Voltage increases were performed on a step-wise basis. Before carrying out the second-dimensional sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), the strips were subjected to a two-step equilibration. The first was an equilibration buffer consisting of 6 M urea, 30% v/v glycerol (BDH Laboratory Supplies, Poole, England), 2% w/v SDS (Merck KGaA, Darmstadt, Germany), 50 mM Tris-HCl (pH 6.8) and 1% w/v DTT. The second step was with a buffer consisting of 6 M urea, 30% v/v glycerol, 2% w/v SDS, 50 mM Tris-HCl (pH 8.8) and 2.5% w/v iodoacetamide (IAA) (Sigma). After the IPG strips were transferred onto the second-dimension SDS-PAGE gel, the strips were sealed in place with 0.75% agarose (USB). SDS-PAGE was performed on 1.0 mm thick 10% and 10% w/v polyacrylamide gels at a constant voltage of 110 V at 10°C using an Amersham Pharmacia Iso-Dalt electrophoresis unit.--

D<sup>4</sup>

--[0105] Silver stained spots were excised manually with a homemade plastic plunger and transferred to a 96-well polypropylene microtiter plate. Each excised spot was washed with 175  $\mu$ L of 25 mM Tris-HCl (pH 8.5) in 50% acetonitrile (Applied Biosystems, Foster City, CA, USA). The plate was sealed with an adhesive film and stored at 4°C for at least 24 hrs. This step was critical for the equilibration of gel spots as it allowed for more efficient enzyme digestion. Prior to the addition of trypsin, the washing solution was replaced with a fresh aliquot of solution and plates were incubated with shaking for 20 mins at 37°C. The washing solution was then aspirated and gel spots were dried in a Savant Automatic Environment SpeedVac AES2010 centrifugal concentrator (Holbrook, NY, USA) for 30 mins. Enzymatic digestion was performed with the addition of 10  $\mu$ L of 0.02  $\mu$ g/L trypsin (Promega Corporation, Madison, WI, USA) in 25 mM ammonium bicarbonate (pH 8.5) (Sigma) to each gel piece and incubated at 37°C overnight with shaking. To enhance peptide extraction, 10  $\mu$ L of 0.1% trifluoroacetic acid (TFA) (Sigma) in

50% acetonitrile was added to each well and the microtiter plate sonicated for 10 mins in an ultrasonic water bath (Crest Ultrasonics, NJ, USA).--

Dy --[0106] Mass analyses were performed according to previously published methods using a PerSeptive Biosystems Voyager-DE STR MALDI-TOF MS (Framingham, MA, USA). In essence, 1  $\mu$ L of the extracted sample from each of the microtitre wells was dispensed onto a MALDI sample plate along with 1  $\mu$ L of matrix solution (10 mg/mL  $\alpha$ -cyano-4-hydroxycinnamic acid (Sigma), 0.1% TFA, 50% acetonitrile). The samples were allowed to dry under ambient conditions. For each sample, the average of 256 spectra was acquired in the delayed extraction and reflector mode. The average of 4 scans (each containing 64 spectra) that passed the accepted criterion of peak intensity was automatically selected and saved. Spectra were automatically calibrated upon acquisition using a two-point calibration with residual porcine trypsin autolytic fragments (842.51 and 2210.10 [M+H<sup>+</sup>] ions). Assignment of peaks was done manually, measured peptide masses were excluded if their masses corresponded to trypsin autodigestion products or to identified proteins adjacent to the spot being analyzed.--

--EXAMPLE 16 - Bioinformatics Findings on HCC-1--

DS --[0126] The Conserved Domain Database (CDD) with Reverse Position Specific BLAST search on the 1-42 amino acids of HCC-1 gave the result as a SAP domain (e-value of 5e-04), which is a putative bi-helical DNA-binding motif predicted to be involved in chromosomal organization and transcriptional regulations (Massari & Murre 2000) found in diverse nuclear proteins. This is supported by PredictProtein where amino acid sequence 197-203 was predicted to contain the nuclear localization signal. There is no predicted trans-membrane segment (using TMAP and PredictProtein), no mitochondrial targeting sequence (PSORT), and no secretory signal (SignalP).--

--[0127] Using PSI-BLAST on non-redundant database, amino acid sequence 1-42 of HCC-1 was matched to vertebrate heterogenous nuclear ribonucleoprotein with identities match of above 45%:

- Heterogenous nuclear ribonucleoprotein U (AF073992) of *Mus musculus* [Expect = 0.005, Identities = 21/42 (50%), Positives = 29/42 (69%)]
- SP120 (D14048) (nuclear scaffold protein that binds the matrix attachment region DNA) of *Rattus norvegicus* [Expect = 0.005, Identities = 21/42 (50%), Positives = 29/42 (69%)]

- ROU\_HUMAN Heterogenous nuclear ribonucleoprotein U (HNRNP U) (Scaffold Attachment Factor A) (SAF-A) (Q00839) of *Homo sapiens* [Expect = 0.012 Identities = 20/42 (47%), Positives = 29/42 (68%)]
- hnRNP U protein (X65488) of *Homo sapiens* [Expect = 0.012, Identities = 20/42 (47%), Positives = 29/42 (68%)]
- Scaffold attachment factor A (AF068847) of *Xenopus laevis* [Expect = 0.021, Identities = 20/37 (54%), Positives = 26/37 (70%)]--

--[0128] Using FASTA3 on SWALL non-redundant database, HCC-1 was matched to various invertebrate translated proteins with E-value below 0.03:

- Q9VHC8 CG8149 protein of *Drosophila melanogaster* [Expect=8e-06]
- Q9N3G0 Hypothetical protein Y53G8AR.d of *Caenorhabditis elegans* [Expect=0.0005]
- Q9LZ08 Hypothetical 22.8 KDA protein of *Arabidopsis thaliana* [Expect=0.021]
- O74871 Conserved hypothetical protein of *Schizosaccharomyces pombe* (Fission yeast) [Expect=0.024]--

--[0129] Physically, this HCC-1 protein may have 2 to 3 domains from coiled-coil and low complexity region predictions:

- PredictProtein Coiled-Coil prediction – the coil is most probably at 30-51 positions. The next possible coiled-coil is at 146-160 positions. Coiled-coil most probably separates the different domains.
- COILS ver 2.2 (Lupas) – at aa 25 – 64 and aa 145 – 172.
- SEG Low Complexity regions predicted 2 regions: at aa 42-79 and aa 165-179.--

---

**IN THE CLAIMS:**

Cancel claims 3-14, without prejudice.